

Sample Questions - Semester IV Data Mining and Business Intelligence

Sr. No.	Question	Option a	Option b	Option c	Option d
1	_____ is a performance management tool that recapitulates an organization's performance from several standpoints on a single page.	Data Cube	data mart	Dashboard	Balanced Score Card
2	Business intelligence works on	Structured semi structured and semi structured data	Structured data	semi structured data	any type of data
3	Integrating BSC with Business Intelligence focuses on	Knowledge , strategy , action	Data, strategy KPIs and analytics		
4	Stochastic Hill Climber is based on the _____ function	Probability	Linear	Non-linear	quadratic
5	Global Optimization Technique satisfies the following conditions	feasible solution	feasible and better than any other solution	better than any other solution	one solution
6	Integrating BSC with Business Intelligence focuses on	Knowledge , strategy , action	Data, strategy KPIs and analytics	Data, information ,knowledge	knowledge and starategy
7	_____ is a system where operations like data extraction, transformation and loading operations are executed.	Data staging	Data integration	ETL	Database management Systems

8	An OLAP tool provides for	Multidimensional analysis	Roll-up and drill-down	Slicing and dicing	Rotation
9	A star schema has what type of relationship between a dimension and fact table?	Many-to-many	One-to-one	One-to-many	many to one
10	The data Warehouse is_____.	read only	read write only	write only	append only
11	Data mining is	The actual discovery phase of a knowledge discovery process	The stage of selecting the right data for a KDD process	A subject-oriented integrated time variant non-volatile collection of data in support of management	A performance management tool that recapitulates an organization's performance
12	KDD (Knowledge Discovery in Databases) is referred to	Non-trivial extraction of implicit previously unknown and potentially useful information from data	Set of columns in a database table that can be used to identify each record within this table uniquely.	collection of interesting and useful patterns in a database	The actual discovery phase of a knowledge discovery process
13	Dimensionality refers to	Cardinality of key values in a star schema	The data that describes the transactions in the fact table	The level of detail of data that is held in the fact table	The level of detail of data that is held in the dimension table
14	_____describes the data contained in the data warehouse	Relational data	Operational data	Informational data	Meta data
15	Suuport of (A) is	Total number of transactions containing A	Number of transactions containing A / Total number of transactions	Number of transactions not containing A / Total number of transactions	Total Number of transactions not containing A

16	How do you calculate Confidence(A → B)?	$\frac{\text{Support}(A \cap B)}{\text{Support}(A)}$	$\frac{\text{Support}(A \cap B)}{\text{Support}(B)}$	$\frac{\text{Support}(A \cup B)}{\text{Support}(A)}$	$\frac{\text{Support}(A \cup B)}{\text{Support}(B)}$
17	What are maximal frequent itemsets?	A frequent itemset whose no super-itemset is frequent	A frequent itemset whose super-itemset is also frequent	A non-frequent itemset whose super-itemset is frequent	A non frequent itemset whose no super-itemset is frequent
18	What is the effect of reducing min confidence criteria on the same?	Number of association rules remains same	Some association rules will add to the current set of association rules	Some association rules will become invalid while others might become a rule	cant say
19	Classification is	A subdivision of a set of examples into a number of classes	A measure of the accuracy, of the classification of a concept that is given by a certain theory	The task of assigning a classification to a set of examples	Forming groups of data based on their similarity
20	Cluster is	Group of similar objects that differ significantly from other objects	Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm	Symbolic representation of facts or ideas from which information can potentially be extracted	A subdivision of a set of examples into a number of classes
21	Considering the K-means algorithm, if points (-1, 3), (-3, 1), and (-2, -1) are the only points which are assigned to the first cluster now, what is the new centroid for this cluster?	(1,1)	(2,-1)	(-2,1)	(-1,1)

22	Which of the following is required by K-means clustering?	defined distance metric	number of clusters	shape of clusters	tree showing how close things are to each other
23	Recall is	the percentage of documents that are relevant to the query and were, in fact, retrieved	the percentage of retrieved documents that are in fact relevant to the query	the percentage of documents that are irrelevant to the query and were, in fact, retrieved	the percentage of retrieved documents that are in fact irrelevant to the query
24	Stop list is	Set of words that are deemed "relevant", even though they may not appear frequently	Set of words that are deemed "relevant", even though they may appear frequently	Set of words that are deemed "irrelevant", even though they not may appear frequently	Set of words that are deemed "irrelevant", even though they may appear frequently
25	Signature file is	a representation of an irrelevant terms that describe the document	a representation of an ordered list of terms that describe the document	a representation of an unordered list of terms that describe the document	a representation of an relevant list of terms that describe the document